Statistical Inference - EDA ToothGrowth

JoySN

January 3, 2019

Analyzing ToothGrowth data in R

Analyzing ToothGrowth dataset for basic exploratory data analysis. The data was explored and checked for different columns values. We tried to find some basic relationships between the columns using plots, confidence Intervals for

Data load and exploratory analysis

```
library(datasets)
library(ggplot2)
data(ToothGrowth)
```

1. Explore the data and columns

```
head(ToothGrowth,2)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
```

```
summary(ToothGrowth)
```

```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000
```

```
table(ToothGrowth$supp) # Supplement
```

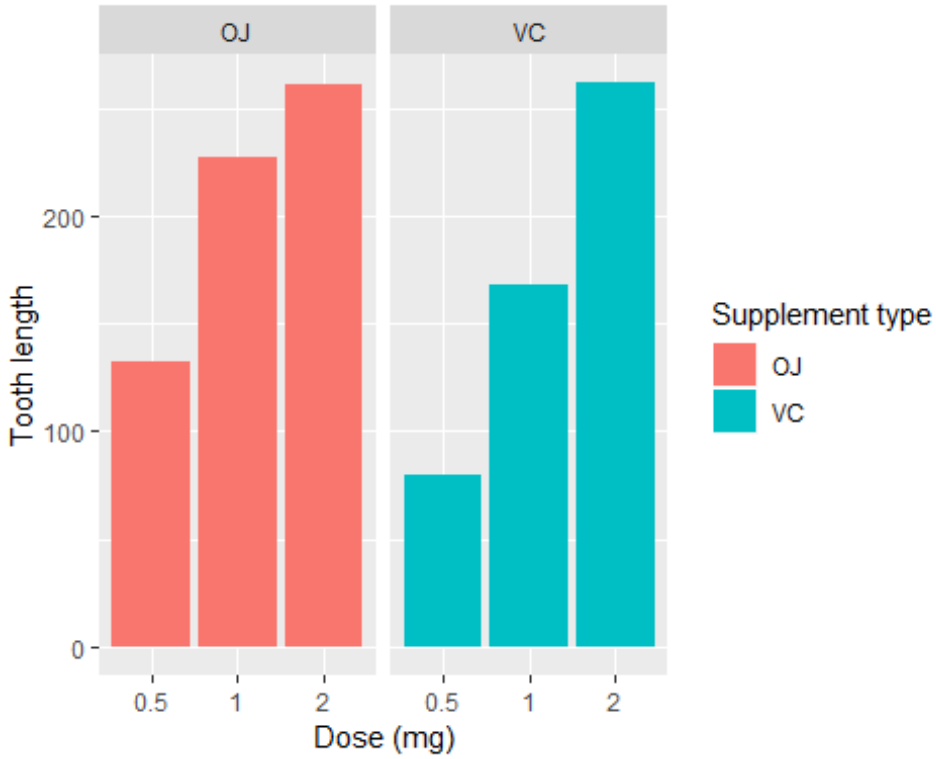
```
## 
## OJ VC 
## 30 30
```

```
table(ToothGrowth$dose) # Dosage
```

```
## 
## 0.5   1   2 
##  20  20  20
```

2. Plot Tooth length vs. Dosage (mg) split by Supplement Type (OJ or VC)

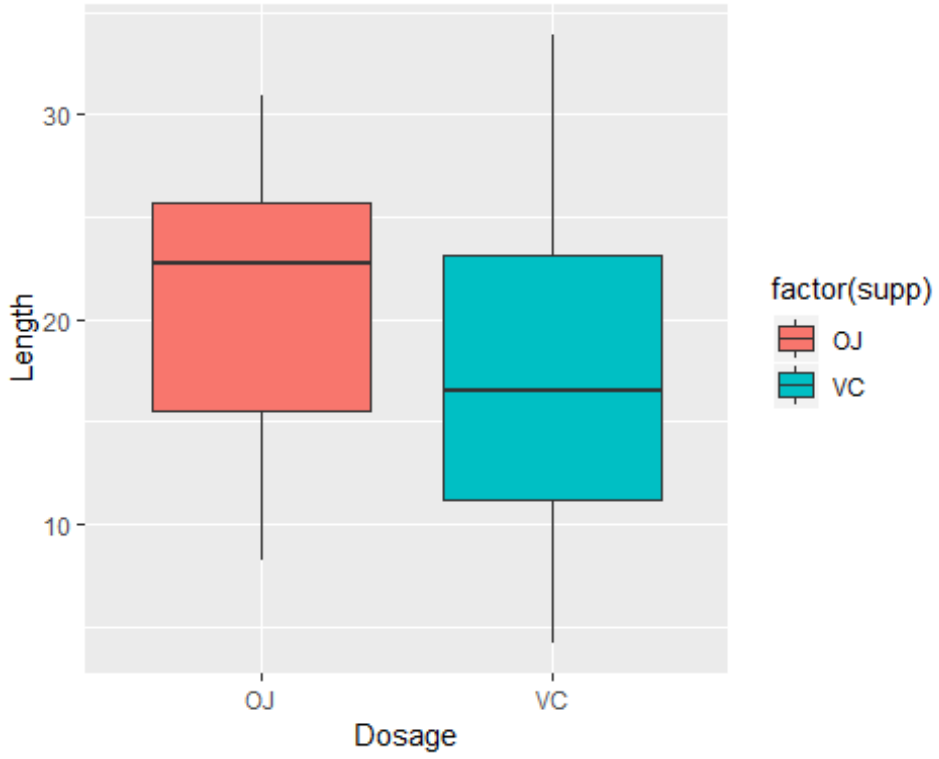
```
gg = ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp))
gg = gg + geom_bar(stat="identity") + facet_grid(. ~ supp) + xlab("Dose
(mg)") +
    ylab("Tooth length") + guides(fill=guide_legend(title="Supplement type"))
print(gg)
```



We can see that - There is a positive relationship between Dosage and Tooth length. - In case of OJ Supplement, 0.5 to 1 cause maximum increase, while for VC Supplement, it is quite uniform accross all dose increase

3. Which supplement has more effect?

```
p1 <- ggplot(ToothGrowth, aes(x =factor(supp), y = len, fill = factor(supp)))
p1 + geom_boxplot() + xlab("Dosage") + ylab("Length")
```



- The plot shows that length of tooth is greater when supplement OJ is given

4. Correlation between Tooth growth and Does

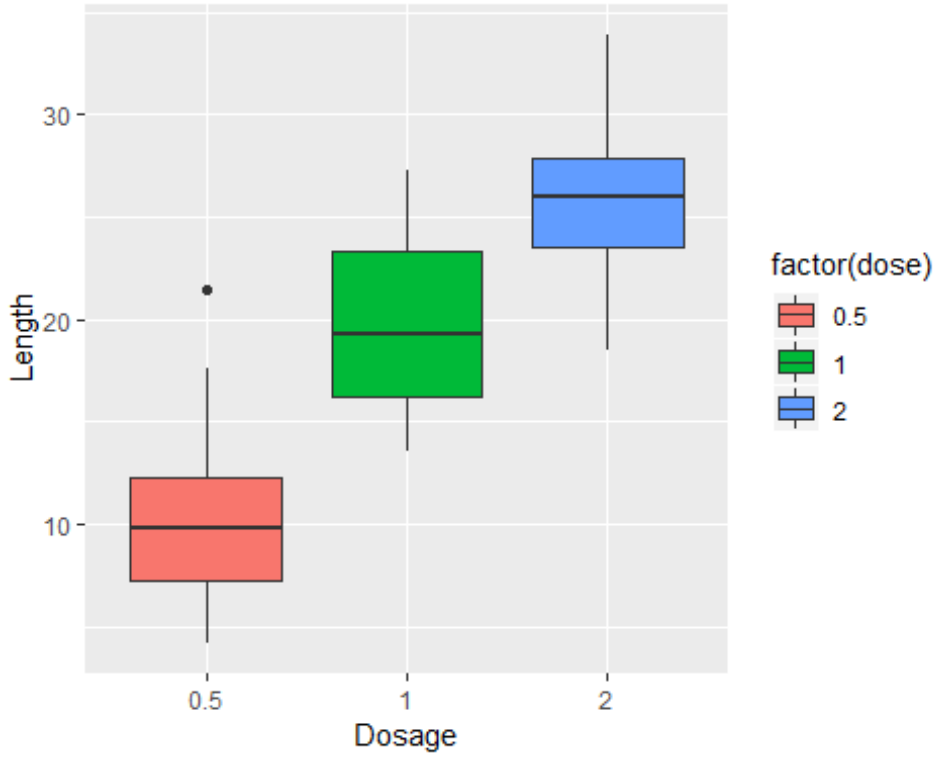
```
cor(ToothGrowth[sapply(ToothGrowth, is.numeric)])
```

```
##             len      dose
## len  1.0000000 0.8026913
## dose 0.8026913 1.0000000
```

- It shows a high correlation between Does and Length of Tooth

5. Plot between Different Does and its effect on Length of Tooth

```
p1 <- ggplot(ToothGrowth, aes(x =factor(dose), y = len, fill = factor(dose)))
p1 + geom_boxplot() + xlab("Dosage") + ylab("Length")
```



- The plot shows more clearly that as the dosage increases, the length of the tooth increases.

Confidence intervals and hypothesis tests to compare tooth growth by supp and dose

6. The 95% Confidence Intervals of Tooth Grow

```
Interval <- (mean(ToothGrowth$len) + c(-1, 1) * qnorm(0.975) *
sd(ToothGrowth$len)/sqrt(length(ToothGrowth$len)))
Interval
```

```
## [1] 16.87783 20.74884
```

7. T-Test for mean difference by supplement type

```
t.test(len~supp, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

- p-value is not small and the confidence interval contains 0, i.e. we cannot reject null hypothesis.

8. T-Test for mean difference by dosage level

```
dose1 <- subset(ToothGrowth,dose %in% c(0.5,1.0))
dose2 <- subset(ToothGrowth,dose %in% c(0.5,2.0))
dose3 <- subset(ToothGrowth,dose %in% c(1.0,2.0))

t.test(len ~dose, data = dose1)

##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735

t.test(len ~dose, data = dose2)

##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100

t.test(len ~dose, data = dose3)

##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

- Each of the above have very small p-values which means the null hypothesis can be rejected.

8. T-Test of Dosage Level

```
dos1 <- subset(ToothGrowth, dose == 0.5)
dos2 <- subset(ToothGrowth, dose == 1)
dos3 <- subset(ToothGrowth, dose == 2)

t.test(len~supp, data=dos1) # Small Dosage = 0.5

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98

t.test(len~supp, data=dos2) # Medium Dosage = 1

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77

t.test(len~supp, data=dos3) # Big Dosage = 2

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

- For Small and Medium dosage, their p-values are small, so we can reject the null hypothesis
- For Big Dosage, we cannot reject the null hypothesis

Conclusion

1. No difference by supplement type i.e. OJ or VC do not have any effect independently on tooth length growth
2. Dosage levels have a significant effect on growth of tooth length
3. For dosage level 0.5 and 1, the OJ Supplement has a higher effect on the length of tooth than the VC Supplement.
4. For dosage level 2, there are so such difference between the supplements

Assumption

1. All samples are IID
2. It is a truly random sample of the population